2019 年 5 月 20 日

“启智、包容、启迪”——iConference 2019 会议综述

■ 巩洪村[1] 沈雨田[1] 曹高辉[1,2]
[1] 华中师范大学信息管理学院 武汉 430079 [2]湖北省数据治理与智能决策研究中心 武汉 430079

摘要：［目的/意义］iConference 是信息科学领域重要的国际会议，文章旨在对 iConference 2019 会议论文进行总结与梳理，形成比较完整的综述内容，为国内该领域研究者提供参考和借鉴。［方法/过程］从研究方法、研究问题、研究过程、研究发现 4 个维度对 iConference 2019 所录用的 24 个主题 77 篇论文进行综述。［结果/结论］iConference 作为 iSchool 的年会，其录用论文主题涵盖了当前信息科学领域理论研究和应用实践的最新成果。从成果可以看出，基于数据驱动的研究范式为信息科学与其他学科的融合发展提供了契机，促使图书馆与信息学科在新的研究场景下发现研究问题；半结构化访谈、问卷调查、扎根理论、统计分析等定性与定量相结合是当前主要的研究方法；研究数据的获取是研究中极为重要的过程。

关键词：iConference 会议综述 数据驱动 信息科学
分类号：G250
DOI：10.13266/j.issn.0252-3116.2019.12.016

1 会议介绍

2019 年 3 月 31 日 -4 月 3 日，2019 年（第十四届）iConference 会议在美国马里兰大学帕克分校举行，本次会议由马里兰大学帕克分校主办，雪城大学和马里兰大学巴尔的摩郡分校协办，会议获得美国国家科学基金会、美国计算研究协会、爱墨瑞得出版社、爱思唯尔出版公司、MDPI 出版社、摩根和克莱普尔出版社、NVIVO 公司、匹兹堡大学计算与信息学院、肯塔基大学信息科学学院、台湾大学图书资讯学系、田纳西大学诺克斯维尔分校信息科学学院等机构赞助。本次会议的主题是“启智、包容、启迪（inform，include，inspire）”，旨在探讨 21 世纪启智的意义，以及如何扩大信息革命的覆盖面，并思考如何更好地启迪个人和组织在快速变化的知识社会中利用信息。全球 60 余位 iSchool 学院院长，近 600 名学者参加此次会议。本次会议共录用 77 篇论文，包括 33 篇 Short Paper（录用率 37.5%）和 44 篇 Full Paper（录用率 33%），录用的论文由 LNCS（Springer's Lecture Notes in Computer Science）会议论文集出版[1]，海报及部分论文在伊利诺伊大学图书馆的 IDEALS（Illinois Digital Environment for Access to Learning and Scholarship）开放获取[2]，以下关于会议主题的综述即是基于这 77 篇会议论文进行的。

2 会议内容

本文对本次会议所录用的 77 篇论文进行综述，以期了解国际范围内本学科最新研究动态和前沿工作。77 篇论文共分为 24 个研究主题，下文将按主题对论文主要内容进行总结。

2.1 科学工作与数据实践（Scientific Work and Data Practices）

数据密集型科学发现作为科学研究的第四范式已在人文社会科学和自然科学的多个学科得到快速发展，高效地共享、利用和管理海量的科学数据是基于数据驱动的科学研究的重要基础保障。肯塔基大学的 Youngseek Kim 基于制度理论和计划行为理论构建研究数据共享理论框架，并采用问卷调查方法采集数据，研究制度压力、资源、个体动机对生物科学家开放共享不同类型的研究数据（基因组序列数据和实验室数据）的影响。其研究发现，制度压力（科研经费资助机构和学术期刊的要求），资源（元数据标准和数据仓储平台）、个体动机（感知职业收益、感知职业风险、个体努力）这 3 个因素对生物科学家开放共享不同类型研究数据的影响有显著差异，建议采用不同的政策措施

作者简介：巩洪村（ORCID：0000-0002-7334-4942），硕士研究生；沈雨田（ORCID：0000-0002-4729-7193），硕士研究生；曹高辉（ORCID：0000-0002-2760-4889），副教授，博士，通讯作者，E-mail：ghcao@mail.ccnu.edu.cn。
收稿日期：2018-05-07 本文起止页码：121-132 本文责任编辑：杜杏叶

鼓励生物科学家开放共享研究数据。劳伦斯伯克利国家实验室的 Drew Paine 和 Lavanya Ramakrishnan 对 5 名天文学和 5 名地球科学领域的数据生产者和数据利用者进行半结构化访谈,了解其科研项目的各个方面,重点聚焦其所用数据及如何获取这些数据,并且询问如何确定用于分析的数据产品的版本和数据产品发生变化带来的影响(预期的和意外的),基于改进的扎根理论研究方法对访谈数据进行编码和分析,探讨科学家在研究工作中是如何认识数据变化的,旨在设计工具和最佳实践以支持和帮助科学家发展和维护他们的研究基础设施,并支持科学家检查数据变化,发现嵌入研究基础设施的关键信息。卡内基梅隆大学的 Ei Pa Pa Pe-Than 和 James D. Herbsleb 以太空望远镜科学研究所(Space Telescope Science Institute)的软件工程师和科学家(黑客马拉松参与者)为研究对象,采用实地考察、半结构化访谈以及问卷调查等混合研究方法,并使用基于 Web 的定性数据分析工具 Dedoose 对访谈记录和观察记录进行开放式编码,旨在研究黑客马拉松如何促进科学软件团队的合作,参与者如何利用活动提供的资源来完成工作、交换知识、扩展社交网络以及培养在日常工作中可以运用的技能和实践能力。

2.2 (大)数据研究中的方法问题(Methodological Concerns in (Big) Data Research)

基于数据驱动的研究范式推动了现代科学研究的发展,同时也还面临着研究方法规范性不足、研究结果可解释性较差、数据分析可视化复杂等问题。围绕大数据研究中的方法问题,武汉大学 Tingting Jiang 等采用扎根理论方法从 14 篇关于信息偶遇的研究论文中抽取了 279 个信息偶遇事件进行描述,形成了 9 个大类、33 个子类、230 个概念标签,抽象构建了信息偶遇过程模型,并对模型中主要元素之间的关系进行详细分析。普林斯顿大学的 Madelyn Rose Sanfilippo 和印第安纳大学的 Chase McCoy 从 Web of Science 中选取了 2010－2018 年在"政府和法律""教育评估"和"信息和图书馆学"三个领域里引用最高的 50 篇论文(仅包括实证研究和明确概述了研究方法的论文),采用元分析方法比较评估数据驱动和统计分析两种研究方法的差异,结果显示传统的统计分析方法更加制度化和一致性,数据驱动的研究方法在信度、效度、可概括性和可解释性方面存在一定的限制。华盛顿大学的 Jaime Snyder 和马里兰大学的 Katie Shilton 从受众、支持结构、许可、用途和可扩展性、图形功能、如何输入或导入数据、图形输出、设计模式、美学、布局和优化功能角度对 2 个可视化工具(R 和 Adobe illustrator)进行比较,并从软件文档(发行说明、用户指南等)、培训材料(在线教程、视频、演示等)、营销和媒体材料(新功能公告、广告、评论、博客文章等)中按时间抽取对软件特性和功能进行描述的内容,以此追踪产品的更新变化。他们同时对这两种工具在专业术语措辞和功能方面的价值进行比较和评价,并进一步讨论数据驱动的工作和数据科学课程在当前环境中的意义,认为在当前环境中,数据可视化实践正在趋同,而软件产品介绍中的专业术语措辞表达仍存在分歧。

2.3 关注"智能"互动和隐私(Concerns About "Smart" Interactions and Privacy)

随着物联网技术的发展,智能设备、智能家居广泛应用于各类生活场景,在给生活带来方便的同时也引发了用户对于隐私泄露的担忧,技术的普适性和隐私保护程度影响着用户对智能设备的接受。奥尔堡大学哥本哈根分校的 Toine Bogers 等采用问卷调查方法调研了 357 名丹麦语用户如何使用智能个人助手,调查结果显示只有 19.9% 的用户使用智能个人助手,大多数用户认为智能个人助手功能不可靠。对 Siri 功能可用性研究发现,在使用丹麦语进行交互的情景下,Siri 存在语音识别能力差、对话反应不自然、无法支持混合语言语音识别等问题。乔治梅森大学的 Chola Chhetri 和 Vivian Genaro Motti 从 Amazon. com 和 Bestbuy. com 网站抓取对 Amazon Echo、Google Home、Samsung SmartThins Hub、Wink Hub 和 Insteon Hub 的用户评论信息 66 656 条,采用手工方式从关注程度(具体、不具体)、情感极性(正面、中性、负面)、时序(采集、传输、存储、分享)、安全原则(保密性、完整性、可用性、可认证性)4 个维度对评论进行编码,分析用户对于智能家居设备的隐私关注点,据此提出提升智能家居产品隐私保护的功能建议。马里兰大学的 Yuting Liao 等和威斯康辛大学密尔沃基分校的 Michael Zimmer 等采用在线调查方法采集了 1 160 个样本数据(包括使用智能个人助手用户和未使用智能个人助手用户),基于社会契约理论和技术接受模型探讨用户使用 IPA 的动机和障碍、对数据隐私和公司遵守与 IPA 数据相关协议的信任担忧是如何影响 IPA 的接受和使用。

2.4 在线社区中的身份问题(Identity Questions in Online Communities)

在线社区的用户通过各种在线行为(如发布信息、购买虚拟装备、设置用户头像等)为自我在社区中构建了一种用户形象,同时,在线社区通过积分机制、等级

机制等方式吸引用户积极构建虚拟身份,虚拟身份的认同是部分用户持续使用社区的重要影响因素。印第安纳大学的 Travis Faas 等使用半结构化访谈的方式调查了 Twitch 直播平台中 8 个拥有 10 个或更少观众的流媒体用户,并利用开放式编码的方式对访谈内容进行主题分析,研究拥有较少观众的流媒体用户持续直播的动机。研究发现,一半受访的用户处于工作、城市或生活方式改变的过渡阶段,他们持续直播的动机有两个:一是工作的需要,在直播的过程中获取观众对其工作的意见和帮助;二是减少孤独感,通过直播可以认识现实生活中不能遇见的朋友。最后,根据研究发现建议流媒体网站为处于生活过渡时期的用户设计直播社区和学习环境,满足用户和观众的学习需求,并为寻求帮助的用户找到能够帮助他们的人。纽约州立大学石溪分校的 Alia Reza 等通过调研与游戏中表现形式和与玩家种族身份的购买行为相关的文献,研究微交易中游戏角色的代表性特征对于不同种族背景玩家购买行为的影响。研究发现,通过微交易购买皮肤可以让玩家选择比默认皮肤更加多样化的外观,这允许玩家实现更大程度的自我个性表达并提升了其社交归属的可能性,但是角色外观(如皮肤的颜色)的商品化可能也会增加部分种族玩家的心理负担。美国西北大学的 Kathryn E. Ringland 采用虚拟民族志方法(virtual ethnographic methods)获取了 Autcraft 中用户 2014 年 5 月到 2017 年 5 月期间约 200 小时的沉浸式观察数据,包括约 5 000 个论坛主题、150 多个博客帖子,并利用迭代和归纳的方式识别、分析和命名数据中出现的现象。其研究发现"自闭症"这个标签不仅是剥夺公民权利的根源,也是赋权和身份认同的来源。他认为,在线社区可以为自闭症患者提供相对封闭和安全的空间,但不能成为治疗自闭症的医学途径;在社区中表明自闭症身份虽然可以获得一种身份的认同感,但也要鼓励自闭症患者勇敢面对现实世界。

2.5 测量和跟踪科学文献(Measuring and Tracking Scientific Literature)

德克萨斯农工大学的 Dong Joon Lee 等从 Scholars @TAMU 中提取德克萨斯 A&M 大学工程学院教师的相关数据,并根据美国国家研究委员会(National Research Council,NRC)所采集的教师获奖和获得荣誉的数据[3]将这些教师分为获得高度声望(HP)奖项、获得著名(P)奖项以及未获得任何奖项 3 种类别,使用 Kruskal-Wallis 检验探索教师研究绩效指标与研究合作实践之间的关系,并利用 Bonferroni 检验确定了研究者类别中绩效指标的相对重要性。研究结果表明,获得 HP 奖项教师的总引文数、引用文章的引用次数、h 指数和出版物总数的平均排名明显高于没有获得奖项的教师;获得 HP 奖项教师的引文数量和出版物总数的平均排名明显高于获得 P 奖项的教师,但获得 P 奖项的教师和没有获得奖项的教师之间没有显著差异。大学管理者和文献计量领域研究者可依据此研究结果将获得奖励与其他研究绩效指标联合使用,同时,还可以利用此研究结果研发机器学习模型预测教师是否有望获得 HP 奖项。罗格斯大学 Michael Lesk 等采集 2014-2018 年期间在 PLoS 收录的文章以及其撤回的 153 篇论文(通过 Retraction Watch 网站),采用统计分析方法研究了开放数据与可信研究之间的关系。结果显示,PLoS 中大约一半的论文都有附加数据,撤回论文中数据可用性的比例低于平均水平,约 27% 的撤回论文有开放数据。研究表明具有高质量、精心策划的开放数据的论文将更加可靠,应该得到更高的可信度,并且有权在大学社区中获得更高的声望,在搜索引擎中获得更高的排名。雪城大学的 Ania Korsunska 以谷歌搜索引擎为数据源,通过超链接引用跟踪某篇特定文章并建立引文网络来探讨科学错误信息的传播过程。具体问题包括:错误信息如何在不同媒体来源之间流动?哪些媒体渠道依赖主要来源,哪些媒体来源依赖二手资源?哪个媒体渠道是错误信息的来源?结果表明,错误信息传播的速度比事实传播更快更广,大众媒体过度依赖二手资料,出版物的声誉会影响公众对其发布信息的可信度,流行媒体出版物互为对方的信息来源。南京大学的 Tong Zeng 等以 PubMed Open Access 中 190 万篇文章的 URL 为研究对象,探讨开放获取出版物中共享资源的可重用性问题。通过逻辑回归方法研究 URL 形式的资源共享是如何随着时间的推移而演变,并探讨影响资源可用性的相关因素。结果表明,URL 已经成为共享资源的主要机制,有一半的链接在 8 年后无法使用,但这种趋势在第二个 8 年期间不会继续,URL 的衰变趋势不遵循物理学研究中发现的典型半衰期分析,资源的使用次数、国际和组织域名后缀以及论文中从属关系的数量与链接的可用性呈正相关,链接的年龄、长度以及印度(in)、欧盟(eu)和中文(cn)域名后缀与链接的可用性呈负相关。

2.6 自动化的不足和功能(Limits and Affordances of Automation)

康斯坦茨大学的 Felix Hamborg 等采用社会科学中已建立的模型和方法来描述和分析媒体偏见,同时

利用计算机科学中可扩展的文本分析方法作为技术支撑,提出一个系统原型(Newsalyze)。该原型通过在一组报道同一事件的新闻文章中选择词汇和标签(WCL)来发现偏见实例,采用美国的 DACA 和 Denuclearization 及英国的 Brexit 新闻证实了该模型的有效性。九州大学的 Emi Ishita 等提出自动化内容分析过程的 3 阶段:文档粒度下确定要标记的文档,在句子粒度下从要标记的文档中确定哪些文本应该被标记,为这些文本标注能够体现人们价值观的标签。从互联网上采集自 2011 年福岛核电站泄露到 2016 年这 5 年间与核能和核电站相关的报道,并由两名研究人员经过 10 轮的标注,最后获得 448 篇报道组成的语料库。然后将 348 篇报道作为训练集对 TinySVM 和 fastText 两种分类器进行了训练,用剩下的 100 篇报道作为测试集对分类器的分类效果进行了测试。同时,由两名人员对其中的 28 篇报道中有价值的句子进行了标注,然后通过比较分类器分类结果的查准率、查全率和 F 值对分类器性能进行评价,结果显示分类器性能较好。牛津大学互联网研究所的 Matt Willis 采用田野调查法对英格兰各地的 6 个初级保健实践点的工作人员进行观察和半结构化访谈,收集了全科医生临床实践工作的相关文档,研究人工智能整合到自动化全科医生临床文献实践中面临的挑战。调查发现,全科医生文献实践工作对及时性要求较高;病历中的笔记由简单、快速到复杂、耗时,差异较大;全科医生在文献实践工作中表现了对人际交往风格和文档记录方式的个人偏好;此外,智能扬声器和其他智能设备能提高医生文档工作效率。研究显示自动化能有效避免全科医生倦怠和压力,然而由于文献工作过程中的隐性信息和环境影响,人工智能无法完全实现自动化地解决复杂任务。因而,作者建议医疗人员和人工智能系统实现协同合作,将更有利于文献实践工作。

2.7 采集弱势群体的数据(Collecting Data about Vulnerable Populations)

对弱势群体(贫困人群、移民、残疾人、语言障碍者等)的信息需求、信息行为研究,有助于公共信息服务机构为其提供有效的信息服务。北卡罗来纳大学教堂山分校的 Jordan Dodson 等采用问卷调查的研究方法,对 83 名美国第 1 代(17 岁之后移民美国)到 2.5 代(出生在美国但父母双方有一方不在美国出生)患有抑郁症的移民大学生进行健康自我跟踪的情况及对抑郁症管理数字工具的使用情况进行调查。结果显示,94% 的抑郁症学生进行健康自我跟踪,但极少有人跟踪心理健康指标,81.9% 的学生不使用 APP 工具,86.7% 的学生不使用其他在线资源来帮助他们治疗抑郁症。该研究提出了为什么不使用健康数字工具以及将来如何使用这些工具的问题。印度尼西亚卡查玛达大学 Safirotu Khoir 等通过分析信息行为领域采用 Photovoice 研究方法的相关文献,从方法论层面探究 Photovoice 在探究移民信息需求方面的优越性及可行性。Photovoice 是一种使用照片传递信息的方法,能够有效弥合研究人员与参与者之间的文化差异及语言障碍,有助于获取参与者的思想和情感。该研究体现了照片在研究弱势群体、非英文背景人群中的研究价值。田纳西大学的 Devendra Potnis 和古吉拉特邦中央大学的 Bhakti Gala 为提高发展中国家穷人的移动应用、金融和信息素养,开发了一款针对穷人的跨学科 6 步工具,该工具包括调查问卷、焦点小组提示及实践培训。该研究为发展中国家公共图书馆、政府促进贫困人群的移动金融交易提供参考。华盛顿大学的 Sara Vannini 等采用小组访谈方法对美国地区参与无证移民工作的 9 名人道主义组织人员的信息隐私意识与做法进行调查,利用定性分析软件 Dedoose 识别主题领域和周期性主题。结果显示,人道主义组织用来收集和存储弱势群体信息的系统存在技术和人为因素风险;未经培训的工作人员和志愿者存在无意泄漏弱势群体敏感信息的情况;缺少负责监管信息泄漏问题的组织。基于此,作者提出人道主义信息活动隐私准则(HIA),以保证人道主义组织在非正常移徙背景下为弱势群体提供充分的隐私保护。

2.8 通过公共图书馆和基础信息设施支持社区(Supporting Communities Through Public Libraries and Infrastructure)

公共图书馆和政府建设的其他信息基础设施对满足社区的信息需求、文化需求等有重要价值,也是社区建设的重要组成部分。肯特州立大学的 Kathleen Campana 和华盛顿大学的 J. Elizabeth Mills、Michelle H. Martin 采用焦点小组、访谈和全国调查等数据收集方法对来自城市、郊区和农村地区的小、中和大型公共图书馆的 157 名图书馆工作人员进行调查,旨在探索图书馆如何收集并利用服务欠缺社区中家庭和儿童的需求数据,从而更有效地为社区成员提供服务。研究结果表明,大多数图书馆都在积极使用各种方法收集数据,对话方法是其使用的主要方法,但对焦点小组、访谈等结构化、持续化的方法使用较少;社区(包括社区成员、合作者、领导者等)是图书馆最常见的数据收集

巩洪村，沈雨田，曹高辉．“启智、包容、启迪”——iConference 2019 会议综述［J］．图书情报工作，2019，63(12)：121－132．

来源，但图书馆也正在寻求关于家庭需求的其他类型数据，例如人口普查数据等；大多数图书馆(76%)都有意识地使用这些需求数据进行服务开发，但只有部分图书馆(35%)做到了持续使用。美国华盛顿大学的 J. Elizabeth Mills 和肯特州立大学的 Kathleen Campana 以 VIEWS2 项目研究为基础，从意向性、互动性和社区三个构建块构建 VIEWS2 概念框架，以支持图书馆中儿童早期学习程序的设计和制作，并与实验组的参与者进行访谈。访谈结果表明，该框架的构建块可能对故事时间提供者的实践产生影响。具体表现在：由于干预(VIEWS2 概念框架)，规划故事时间的早期识字内容的意向性得到增加；内容互动性有助于帮助他们扩展并探索新的活动；与同事之间的互动能够减少孤独感并获得经验和工作反馈。美国佛罗里达州立大学的 Sang Hoo Oh 和 Marcia A. Mardis 通过文献分析法考察了美国农村宽带能够在多大程度上支持先进制造业的发展，具体研究包括：先进制造业对农村发展的重要性、农村地区的宽带情况、农村宽带互联网对先进制造业重要性及相关政策文件。研究结果表明，国家采取了各种举措和计划来支持农村地区先进制造业的发展和宽带服务，但仍有部分农村地区缺乏宽带接入，且获得的宽带互联网服务质量较差；农村社区可以从高技术创新产业的涌入中获益，但目前并不清楚这些社区能否提供宽带支持。美国西蒙斯大学的 Colin Rhinesmith 等为公共图书馆获得有关其宽带互联网连接的速度和服务质量数据，利用参与式研究方法与美国十家公共图书馆的研究小组共同设计宽带测量平台和培训手册并提出了一个用于设计、开发和实施美国公共图书馆的开源宽带测量平台的信息系统设计方案，以协助公共图书馆更好地了解其网络基础设施与数字服务之间的关系以及为美国的地方、州和国家宽带规划工作提供信息。

2.9 学术环境中的信息行为(Information Behaviors in Academic Environments)

大学生是信息素养较高、信息需求差异较大、信息利用频率较高的群体，对该群体的信息行为进行研究具有重要的应用价值，同时，由于大学生研究样本数据更容易获取，对其信息行为的研究一直是信息行为研究领域的重要议题。萨福克大学的 Vincent Grimaldi 等采用半结构访谈的方式调查了 3 个美国研究机构 15 名大学生运动员，研究信息与通信技术(ICT)对大学生运动员从学生过渡到毕业生期间在信息搜寻、日常行为和长期目标制定中的作用。研究发现 ICT 能够提高大学生运动员的信息搜寻能力和信息获取效率；使用 ICT 已经成为大学生运动员的日常行为，使用 ICT 能够增强肌肉记忆、优化比赛策略并进行针对性训练；最后，ICT 对于大学生制定长期专业发展的目标具有积极作用。西蒙斯大学的 Kyong Eun Oh 和 Rong Tang 通过移动应用程序的日记模板收集了美国 51 名大学生 3 周内使用手机处理新闻的日记条目(每条日记中包含参与者对新闻的评价和情感)，采用日记研究(diary study)的方法研究大学生如何通过手机浏览新闻及其对包容性/排他性的政治社会新闻的反应，发现大学生更喜欢包容性的新闻，而拒绝排他性或不公平性的新闻，表明美国大学生对社会正义和政治问题有敏锐的认识。

2.10 数据驱动的叙事和建模(Data-Driven Storytelling and Modeling)

叙事(storytelling)是近年来信息科学领域的重要研究内容，是一种以事件为核心的信息组织与信息展示方式。宾夕法尼亚大学的 Xiying Wang 等人以参与当地艺术节的人为研究对象，开发了 Web 应用程序 Memofest 收集人们的节日故事，并采用半结构化访谈的方法对 18 位参与者进行了访谈及编码。从节日故事中确定了 4 种主题：节日活动、艺术作品、社交活动、回忆地点的变化，从访谈内容中确定了 3 种拥有平行体验的参与者类型：体验探索者、怀旧旅行者和熟悉探索者，最后针对 3 种参与者给出了信息技术如何促进地方发展的相关建议。马里兰大学的 Zhenpeng Zhao 等为探讨漫画叙事中分区和序列(帧的排序形成叙事)两种叙事机制的重要性，以数据漫画为研究对象，采用实验法，让参与者利用复杂的信息图表来回答给定问题，与用相同的信息图表组织成的数据漫画来回答问题进行比较，让参与者从吸引力、速度、空间效率、易用性和趣味性进行评分，研究结果表明数据漫画提供了一种比单一可视化更有效的讲故事方式，分区和序列机制有助于参与者聚焦和回忆故事。中山大学大学的 Enguo Zhou 等为深入研究创新扩散的过程，提出了一个新模型——ABGM(基于采纳行为的图形化模型)，该模型从采纳行为的角度对创新扩散过程进行了建模。该模型定义了两种相关关系：同质性和异质性，并在统一的框架内将社交网络、创新网络和相关网络进行有机联系。接着，该模型在 2006－2015 年的大型人工智能出版数据集上得以实验，实验结果显示，与现有的 3 种基线(逻辑回归、支持向量机、Adamic/Adar)进行比较，采纳行为之间的同质性和异质性为预测采

纳决策提供了相关信息,个体用户采用创新的概率受相关网络扩散过程的影响显著。

2.11 在线活动(Online Activism)

宾夕法尼亚州立大学的 Shipi Dhanorkar 和 Mary Beth Rosson 以众包的视角审视网络信访平台,认为这些平台可以看作是专门为推动社会变革而设计的一种特殊的众包系统。采用目的性抽样方法,选择 Change. org,Care2 petitions,MovenOn. org 和 We the people 四个网络信访平台,从发起请愿书,更新请愿书,签署请愿书和请愿书分析4个维度进行了对比分析,并以此为基础提出了网络信访平台设计改进的思路:提供参与的动机、为请愿发起者建立声誉、为集体行动实现临界质量和消息框架,以平衡在线请愿的"开放性"和"质量"。加州大学的 Jennifer Pierre 采用混合研究方法对 Facebook5 个页面上的社交媒体网站帖子从总体情况,点赞、评论和分享行为,媒介,主题4个方面进行内容分析,重点是原始文章的内容以及相关的评论和活动。旨在考察表达或意图唤起 kama muta 的内容流行程度,探讨 kama muta 在激发个人职位和更广泛的社会参与社会媒体活动中的作用。研究结果表明 kama muta 诱导的内容将被更广泛地分享和喜爱。牛津大学的 Pu Yan 采用案例研究的方法,利用微博社交平台,从"2014年2月东莞扫黄事件"中抽取5个政府/媒体账号,5个知识分子/网络活跃分子的对话数据进行可视化分析,研究表明,此次事件中官方与民间社会不仅存在争议,而且形成了两个具有不同类型特征的群体。值得注意的是,政府话语网络显示出更紧密的群体内部联系。相比之下,民间社会无论是线上还是线下都缺乏足够的协调与合作。中介人(微博)为意见交流提供了渠道,通过易感染(SI)模拟,论证了中介人(微博)在会话网络信息扩散中的结构性作用。

2.12 数字图书馆,数字监管和永久保存(Digital Libraries, Curation and Preservation)

丹佛大学的 Peter Organisciak 等使用算法识别 HathiTrust 数字图书馆中不同粒度级别的重复作品,包括从现有版本的变体内容(相同的工作,不同的表达)到相同内容的变体演示(相同的表达,不同的表现形式),再到不同的印刷副本(相同的表现形式,不同的项目)。案例分析显示,在大型数字图书馆中重复工作调整具有极大的复杂性与挑战性。德克萨斯大学奥斯汀分校的 Ayse Gursoy 等通过阅读2009年至今的200多个英雄联盟数字游戏的补丁说明文档,分析了英雄联盟发展过程中记录的各种角色变化以及随着时间的推移这些变化产生的影响。案例分析结果显示,在动态复杂的数字对象中,变化可以具有不同的功能。了解复杂的数字对象如何随着时间的推移而发展,以及这些对象如何被构造为更改对象,可以为开发出更细微的模型提供借鉴。加州大学欧文分校的 Benedict Salazar Olgado 提出一种将视频游戏保存理解为一个新生领域的布尔迪厄夫方式,其话语和实践由本体论差异和冲突的权力结构塑造。这可以通过一系列评估来说明,这些评估一方面将视频游戏视为物质一样的永久存在,另一方面又认为电子游戏具有短暂性。视频游戏保存和倡议可能倾向于这些极端中的一个,其重点是确保游戏的可玩性或记录游戏玩法。通过对参与者进行分类,可以进一步了解这些相互竞争的价值观,从而说明其中权力关系。作者最后指出视频游戏保存领域的关键取决于下一代视频游戏档案管理员是否能形成具有明确共同价值观和理念的社区。

2.13 社交媒体文本挖掘和情感分析(Social-Media Text Mining and Sentiment Analysis)

马里兰大学的 Brooke Auxier 等人以 Twitter 上超过10K粉丝的92位个体影响者账户为研究对象,使用 VADER 情绪分析包和心理语言学文本分析工具 LIWC 分别对健康影响者推文的情感分布规律和语言使用风格进行分析。研究发现,健康相关主题中涉及最多的是促销性主题,其次分别是健康/运动信息、激励、自我披露、体育相关等。在情绪使用上,男性和女性影响者的推文中使用的负面情绪没有统计学上的显著差异,但女性影响者推文包含的积极情绪高于男性影响者推文;与一般健身账户相比,自我认定为健身"训练者"或"教练"的人之间使用的积极或消极情绪的数量没有统计学意义。武汉大学的 Sijing Chen 以北德克萨斯大学图书馆提供的 The Hurricane Harvey Twitter 数据集为研究对象,利用 SentiStrength 算法探究了灾难情况下社交媒体用户的情绪、位置和时间之间相互作用,及其对推文转发模式的影响。锡拉丘兹大学的 Radhika Garg 对社交平台上帖子及其跟帖的情感分布进行分析,探究社交平台 Reddit 上的讨论对用户决定使用或放弃可穿戴设备的影响。该研究基于2015年12月至2017年12月期间在 Reddit 子版块 r/androidwear 上发表的6 680篇帖子和相应的50 867条评论进行情感分析发现,在 Reddit 用户对帖子的响应方面,帖子的情绪对相应评论的情绪有显著影响,大多数针对正面和中性帖子的评论都是正面的,即使很多带有负面情绪的评论在回应负面帖子时出现,这些帖子也会看到事实

驱动的中立讨论。此外，帖子和评论的情绪将影响用户决定使用或放弃设备，即当其他 Reddit 用户提供解决方案时，或者讨论收到大多数积极的或事实驱动的中立评论时，人们决定继续使用他们的设备。南加州大学的 Amir Karami 采用情感分析和主题建模技术，结合定性编码，提出了流行度分析框架，用于发现政治家受欢迎原因。她采集并分析了 Twitter 上 450 万条与美国政治家、参议员伯尼・桑德斯有关的推文，使用 LIWC 工具得到其中正面情绪的推文，并利用 LDA 模型对正面推文进行主题分析，得到了 10 个经济维度下的主题分布情况，发现桑德斯受欢迎的 8 个经济原因，分别是大学学费、工作收入、福利待遇、就业和收入、最低工资、货币政策与美联储、权利、移民。

2.14 公共领域的数据和信息(Data and Information in the Public Sphere)

香港城市大学的 Yong Ming Kow 等人对通过滚雪球采样方法招募到的 21 名香港居民进行了基于任务的信息搜索和深度访谈的半结构式访谈(12 男，9 女，20－62 岁)，以了解居民获取新闻信息的首选来源和搜索路径，并通过扎根理论对访谈信息编码后进行迭代归纳分析，以探索居民对新闻中虚假错误信息的处理方式。研究结果表明，大多数公民都能意识到错误信息的存在，但只有小部分人会寻找更多信息源与错误信息进行交叉对比，新闻阅读中存在自我理性，居民倾向于阅读简短且来自权威机构的新闻。研究建议交互设计人员除了进一步提供更多丰富的新闻资讯来辅助消除新闻中的错误信息外，更应该考虑长期解决方案，例如加强市民新闻素养的培养等。都柏林大学的 Kalpana Shankar 等对近年来家谱数据供求服务的公私合作关系的相关研究进行了文献梳理，旨在探讨公私合作关系在数据服务领域面临的挑战和发展趋势。研究发现，家谱数据服务中公私合作关系的产生原因是图书馆和档案馆拥有大量私企和历史研究者感兴趣的家谱资料，但缺乏将资源广泛传播的能力；家谱数据服务的公私合作过程中存在紧张关系和访问差异；在合作中制定的服务条款包括排他性协议和保密协议；数据服务在进行公私合作前需要大量规划思考，考虑管理、存储、访问的成本与收益问题，发挥数据价值。武汉大学 Di Wang 等人以中国公民为研究对象，通过在线问卷调查获取用户的社会人口信息、公民对政府开放数据的了解程度、公民对政府开放数据的需求主题等信息，用李克特量表量化用户对政府开放数据的需求程度，通过数据分析和检验证明了三个假设(用户、数据、使用类型与利用率之间会相互影响)，旨在为政府提供开放数据、用户和使用情况这三者间的关系以改进服务。研究结果表明，目前了解和使用政府公开数据门户网站的人数有限；拥有不同社会人口特征的用户对开放数据中的各个主题表现出不同程度的需求；日常生活信息和反腐败是公民利用政府开放数据的 2 种主要类型。研究建议政府未来需要提高公民对开放数据网站的认识，提高数据集质量，以刺激公民对政府开放数据的使用意愿。

2.15 多媒体内容(Engaging with Multi-media Content)

奥尔堡大学哥本哈根分校的 Toine Bogers 等从 Reddit 网站上的/r/gamingsuggestions，/r/gamesuggestions，/r/tipofmyjoystick 3 个组中抓取了 2 266 个关于电子游戏的帖子，并从中筛选出 521 个表达电子游戏信息需求的帖子作为研究数据，采用开放性编码、主轴编码等方法进行编码，归纳总结出用户关于电子游戏信息需求表达的 7 个一级要素和 41 个二级要素，统计分析发现内容、元数据、体验、情境和交互性是用户在进行游戏信息搜索中常提到的 5 个要素，其中内容和元数据是用户在搜索已知内容特征的游戏时经常使用的描述要素，而体验、情境和交互性是用户在搜索相似的游戏时经常使用的描述要素。匹兹堡大学 Kayla M Booth 等选择运动和营养这两个健康主题，从平台、信息源等外部线索和基于视频自身的视觉和内容交互因素两个方面对美国 30 名 13－18 岁青少年进行半结构化访谈，探究美国青少年是如何评价社交媒体中提供的具有高风险的视频类健康信息。研究发现，在外部线索中，平台提供的广告和对这些视频的评论，特别是消极、夸大的评论会对青少年评价视频类健康信息产生负面影响，而详实、有趣的视频内容则会在青少年群体中得到更高的评价。视频信息的图像质量和视频的一些定量指标，如点赞数、浏览量等会对青少年评价社交媒体中的视频类健康信息产生影响。而在观看过程中，视频中出现的人物的能力和业务水平同样会对青少年评价视频类健康信息产生影响。新西兰怀卡托大学的 Sally Jo Cunningham 等对 2002－2016 年间怀卡托大学所有关于音乐行为研究的访谈数据进行分析，探究流服务会不会带来个人音乐收藏的转变。研究发现个人音乐收藏主要分为 3 个阶段：2002－2004 年的实物收藏和数字收藏混合阶段，2006－2007 年向数字收藏转变的阶段，一直到 2016 年大规模的音乐集合和音乐流媒体阶段。第三阶段主要以数字形式收藏音乐，

相比于前两个阶段,该阶段用户之间的界限十分模糊,而且用户很难评估音乐收藏规模的大小。该阶段也可以方便用户随时收藏音乐,并且用户不需要担心转移成本的问题。总结发现,目前的流服务确实让用户收藏音乐的方式发生了转变。

2.16 理解在线行为和体验(Understanding Online Behaviors and Experiences)

武汉大学的 Tingting Jiang 等针对某网站两个月的服务器日志文件中的 39 990 200 个用户点击流数据,结合网站首页抓取的新闻新颖度和流行度进行分析,采用 Mann-Whitney U 方法检验新闻新颖度和流行度与用户点击行为之间的关系,旨在调查新闻时效性和受欢迎程度对用户点击行为的影响。通过分析点击量数据发现,这两个指标对新闻标题的选择产生了影响,同时,"从众效应"表明,当很多人也相信信息时,个人倾向于相信它。这一结果对新闻提供者在创造有效的新闻标题和更负责任地出版和传播新闻方面具有重要意义。武汉大学的 Dan Wu 选取"数字敦煌"平台作为案例,采用两种数据采集方法,一种是 Tobii Studio 记录的眼动追踪数据,对 40 名参与者(其中男性 20 名,女性 20 名,来自 19 所学校的 32 个专业,学历平均分布于大一到研究生)进行眼动实验数据收集,眼动特征主要选择注视点、眼跳及其派生指标,实验任务包括自由浏览、可视化搜索和交互;另一种是问卷调查,是在系统可用性量表(SUS)的基础上改进的量表,每个任务都有相应的问卷,选取了 31 名凝视采样率达到 85% 以上的受试者进行分析。其研究旨在将用户的注视特征与体验特征相结合,构建一个评估模型,对数字人文平台的可用性进行测试,并找出它们之间的关系。结果显示注视次数越多,注视总时间越长,用户对 AOI 的兴趣越高,说明关注程度越高;跳码数少,跳码总距离小,跳码时间间隔短,表示用户视觉搜索次数少,容易找到目标,平台结构和布局让用户满意,效率高;用户满意度对平均注视时间、平均眼跳距离无显著影响。台湾师范大学的 Hsin-Kai Yu 和 I-Chin Wu 提出一种搜索模式分析方法,邀请 12 名用户参与实验,在以目标为导向或以探索为基础的不同需求状态下,使用 Morae 来记录参与者的搜索和在线购物行为,通过执行模拟购物任务实验,观察不同的推荐功能如何帮助不同类型的用户进行在线购物,旨在总结不同需求状态下的用户搜索路径和模式,帮助电子商店设计网站。

2.17 算法(Algorithms at Work)

宾夕法尼亚大学的 Jiawei Chen 等通过实验法和问卷调查对 23 名美国大学生进行技术跟踪调查(为被试者手机安装 app 采集数据),研究在不同的情境感知场景下用户使用交通共享应用程序(请求搭车或作为司机帮助别人)的行为、态度和需求,旨在探索影响社区协同生产的因素以改进推荐算法。采用聚类和模拟分析发现除了传统的时间、地点、人物关系等因素外,一些社区参数和时空背景如用户资源、情绪、能力、车辆尺寸等都是影响用户进行社区协同生产的重要因素。研究结果表明在推荐算法缺失情境参数时可以将社区参数和时空背景因素作为替代补充。北卡罗莱纳大学教堂山分校的 Mohammad Hossein Jarrahi 等人通过视频访谈、政策文件分析和归纳分析对线上自由职业平台 Upwork 上 33 名年龄在 20 – 58 岁间的员工、平台上 26 份政策文件和 98 个论坛主题进行实证研究,探索算法管理(为平台员工自动分配不同的资源任务)是否会带来信息不对称,旨在了解在线数字劳动平台上的员工对平台算法的理解和利用情况。研究结果表明,平台的算法规则会对在线工作者的等级评分、检索排名、薪酬报价等产生影响,产生的信息不对称会影响在线员工对工作的掌控;但在线员工并不完全是算法的被动接受者,拥有算法管理能力可以使用户感知、规避甚至操作算法;算法管理是一个社会技术过程,是平台和用户之间持续交互的结果;理解和利用算法能力的素养可以被视为全球市场中个人竞争优势的核心。威斯康辛大学麦迪逊分校的 Alan Rubel 等人通过对"2017 年 Facebook 反犹太人自动广告投放事件"和"斯康辛州最高法院法官使用 COMPAS 专利风险评估工具编写犯罪嫌疑人调查报告事件"两个案例进行分析,提出"机构洗钱"的概念以比喻组织机构使用算法决策系统时将全部后果责任转移给算法本身从而逃避道德规范约束、模糊自身责任的错误行为,并解释了"机构洗钱"与"掩饰行为"和"责任缺失"的关系。研究结果表明,用算法系统进行重要决策存在潜在的道德风险。

2.18 技术社区中的创新和专业化(Innovation and Professionalization in Technology Communities)

南京科技大学的 Yao Cai 等以 2006 – 2015 年收录于 Web of Science 中的 86 种情报学和图书馆学期刊上发表的共 16 732 篇文献为数据集,对关键词进行共词分析,揭示 iSchool 研究主题以及主题的演变过程,旨在探讨 iSchool 的发展是否依然围绕着"信息,技术和人"展开。研究确定了"图书馆""信任""知识管理""社交网络""文献计量学""信息检索""信息技术""文本挖掘"和"信息素养"9 个具有跨学科特征的研究

主题。结果表明,对信息的研究始终是 iSchool 的发展主线,技术是促进 iSchool 发展的重要因素,为用户提供更好的服务是 iSchool 发展的最终目标。克莱姆森大学的 Guo Freeman 等人通过对业余独立游戏开发者的 12 次平均时长为 80 分钟的半结构化深度访谈,获取关于独立游戏开发者如何进行创新的信息,并对访谈内容进行迭代编码。旨在通过研究独立游戏开发者的创新模式,探索一个由分布式个体组成的非正式集体学习网络,构建一种可以自下而上参与创造性技术实践的创新生态,揭示创新生态学的运作机制。研究结果表明,解决各个领域的问题会激励开发者进行创新;来自社区的帮助、交流与鼓励支持了创新;政治与政策因素会对创新起到促进/阻碍作用。IBM 研究所的 Christine T. Wolf 对一家大型全球技术公司的 11 名机器学习开发人员进行半结构化访谈研究,调查开发人员的日常工作实践以及工作中使用的产品等信息。运用归纳和主题分析技术,探索开发人员对机器学习领域产生兴趣过程、进入该领域时遇到的障碍以及他们在这些过程中使用的信息模式。旨在探索信息在职业选择与职业认同中的作用,信息使用与职业认同间的关系。研究结果表明,不同类型的信息激发了个人对机器学习的兴趣;在工作中从与同事交流和在线平台上获取的信息可以帮助开发者解决工作中的问题;信息的使用作为一种组织原则,它可以将一个专业群体联系在一起。

2.19 Twitter 上的信息行为(Information Behaviors on Twitter)

伊利诺伊大学的 Aseel Addawood 等从 Twitter 采集了 2016 年 1 月 1 日至 2016 年 11 月 28 日期间两个健康主题(MMR 疫苗和健康饮食)的 146 403 条相关推文,根据互斥域词典将 URL 分为新闻、视频、虚假新闻、社交媒体、博客、商业、政府、科学、教育机构、健康杂志、国家或州立专业医学协会、健康保险 12 个类别,根据 Klout 分数(衡量用户整体影响力的指标,前 25% 被认为有影响力),在个人和组织层面识别有影响力的 Twitter 用户并分为不同类别。通过分析相关用户消息的 URL 类别、用户类别及其关系,旨在探讨不同用户类别与其 URL 共享模式之间的关系。研究表明不同类型有影响力的用户会参与不同的健康主题,影响力不是通过共享科学或事实证据获得的,而是可以通过个人关系获得,如当涉及有影响力的用户分享虚假新闻时(如讨论疫苗相关主题的用户分享假新闻 URL 的可能性是讨论健康饮食的两倍多)。纽约州立大学布法罗分校的 Dan Albertson 收集了 2018 年 ACM/IEEE 数字图书馆联合会议(JCDL)、加拿大信息科学协会年会(CAIS)和数字时代图书馆会议(LIDA)在会议之前、期间、之后的推文和元数据,并将推文中的数据元素作为变量(原始推文、转发推文、原始推文和转发、喜欢、喜欢和 RT 计数的总和、转发的原始推文数量、原始推文,RT 和类似计数的总和、不同的用户 ID)进行不同时间点或天数的比较分析,旨在报告最近 LIS 研究型会议的社交媒体使用和活动情况,提供描述性的结论用于分析未来学术会议中的社交媒体活动。德克萨斯大学奥斯汀分校的 Nitin Verma 等通过实验和问卷调查的方法研究人口统计学特征(性别、年龄、教育程度、政治倾向、社交媒体使用情况)和人类价值观动机类型(成就、慈善、遵从、享乐主义、权力、安全、自我定向、刺激、传统、普通性)对社交媒体用户被观察到的信任行为、自我陈述的信任行为和实际信息行为的影响。研究表明,自由主义者更倾向于信任主流媒体和科学期刊,页面中超链接的内容影响用户信任。高安全需求的用户更信任主流媒体,并且容易关注到超链接及点击虚假新闻。人口统计学特征和价值观都可以用来预测用户的在线信任和信息行为,同时,价值观在塑造人们与社交媒体帖子的超链接及在线信息界面的其他特征的互动中有重要作用。

2.20 数据挖掘与自然语言处理(Data Mining and NLP)

印第安纳大学伯明顿分校的 Zheng Gao 等以 IMDB 提供的 2016 年 10 月之前制作的电影为数据源,以 IMDB 用户平均评分作为衡量批判绩效的指标,用其投资回报来表示财务绩效,定量分析了电影元数据的基本特征,以及综合计算的复杂电影特征,并利用支持向量机(SVM)根据识别的特征预测电影是否成功。结果解释了电影产业中的三种常见现象:在所有类型的电影中,家庭剧更容易吸引观众;一部电影的成功很大程度上取决于其演员过去事业的成功;导演和演员之间的稳定合作更有可能取得长期的电影成功,尤其是在系列电影中。锡拉丘兹大学的 Yingya Li 和 Bei Yu 以生物医学文摘中结论部分的发现句为研究对象,从 PubMed 中抽取了 1 100 个包含 5 个常见健康主题的生物医学摘要进行随机临床试验研究设计分组,其中 1 000篇作为训练集,100 篇作为测试集,从精准率、召回率以及宏观平均 F1 得分等指标比较了基于规则的方法以及基于机器学习的方法的性能,实验结果表明基于规则的方法和基于词袋的机器学习方法都具有高

精度,通过简单的基于规则的方法能取得更好的性能表明,尽管先进的机器学习方法可以捕获主要模式,人工专家在这样的专门任务上仍然可能表现得更好。锡拉丘兹大学的 Lu Xiao 和 Xin Huo 以罗格斯大学的参数挖掘小组提供的开放存取语料库为数据源,选取了其中的 271 条包含理由的文本内容,共1 124个语句,并标注了具有权威声明的文本,研究在网络传播内容中理由与权威声明之间是否存在相关性。结果表明包含权威声明的语句为 42 个,仅占总量 3.7% ,说明当用户在这些活动中向其他人展示理由时,几乎没有试图增强他们的可信度。

2.21 基于线下经验的信息产品设计(Informing Technology Design Through Offline Experiences)

匹兹堡大学的 Julia Cope 等对加利福尼亚州、宾夕法尼亚州和马里兰州三个地区的 12 名消防员进行半结构化访谈,获取关于消防员救援中使用导航的培训过程、救援任务中使用的导航策略、遇到的问题以及解决方案的数据资料,基于扎根理论对访谈信息进行编码分析,旨在探索消防员在紧急救援中如何对周围的信息做出判断并制定空间搜索及救援行动策略,并将结果纳入消防员个性化导航辅助工具的设计中。研究结果表明消防员主要基于 4 个目的使用各种导航策略:在救援中快速建立搜索路径、改善视力可见度和可见范围、通过对建筑物内部结构的演绎创建认知地图、在救援中做出定向(确定方向)决策。北京大学 Dede ma 和 Pengyi Zhang 采用在线问卷调查方法调查了中国 4 个城市 1 129 位乘车共享应用程序用户,研究使用乘车共享应用程序时乘客的情绪体验(快乐、满足、惊讶、忧虑、失望、愤怒)以及情绪体验与用户特征(性别、年龄、学历、使用频率/周、支出/周)、使用环境(时间、舒适度、导航精准度、路况、紧急程度、是否有其他乘客)与应用程序的交互满意度的关系,旨在对乘客的情感体验提供初步了解,改善此类应用程序的设计。研究表明使用乘车共享移动应用程序时,乘客感受到的积极情绪比消极情绪多;不同消极情绪体验的用户有不同的情感发泄路径和行动倾向;使用背景、互动满意度和用户特征与乘客的情绪体验有关,当与乘车共享应用的互动更成功时,乘客会获得更积极的情绪,女性和使用频繁的用户有更多的负面情绪。费城大学的 Swathi Jagannath 等在美国大西洋中部地区的儿科教学医院观察并视频记录了 44 次使用电子流程图记录医疗复苏文档的过程、遇到的挑战以及解决方法,并对 24 名护士记录员进行了访谈,旨在了解护士记录员的信息行为及其对医疗复苏期间使用新实施的电子流程图的态度,探索在高风险医疗环境中从纸质到电子文档过渡过程的信息行为和态度差别,改进电子流程图以支持其设计指南在复苏期间使用。研究表明复杂医疗环境中的电子文档在支持患者持续护理方面具有优势,但在流程图导航、复苏数据自动填充、实时文档创建的使用和实践方面面临挑战。

2.22 用于健康管理的数字工具(Digital Tools for Health Management)

加利福尼亚大学 Elizabeth V. Eikey 等通过 EDE - Q 6.0 等 4 份问卷评估了 24 名在校女大学生的饮食紊乱症(EDs)的患病程度,并通过半结构化访谈的方式对 24 名女大学生进行访谈,研究她们是否向其他病患推荐一般意义上的饮食和健康 APP 用于恢复病症以及她们是否会使用专业的 EDs 治疗 APP。研究发现,大部分用户会使用一般意义上的饮食和健康 APP,但是她们并不会将这些 APP 推荐给其他病患,而且大多数也不会使用专业的 EDs 治疗 APP。通过对访谈数据的编码,总结发现用户不知道其存在是导致用户不使用专业的 EDs 治疗 APP 的主要原因,此外还有不受欢迎且不熟悉、不愿意使用和这些 APP 可用性差等原因。宾夕法尼亚州立大学的 Jomara Sandbulte 通过对 21 名成年人以及 16 名老年人的焦点小组访谈,探讨和研究促使家庭中不同的几代人之间持续共享健康信息的因素有哪些。研究发现,与亲人之间的亲密关系会很自然地让亲人之间持续共享健康信息;如果身体健康的话,持续共享健康信息有时候可能会让家庭成员之间感到有些麻烦。不过当身体出现不适的时候,这种健康信息的共享就显得尤为重要。本研究最突出的一个发现是转折点会激励和促使家庭成员之间持续共享健康信息。转折点指的是家庭成员中突然发生重大的健康事件,如突患某种重症。此外,家庭遗传史和开放的家庭环境也会促使家庭成员之间持续共享健康信息。卡内基梅隆大学的 Shikun Zhang 等在 MTurk 平台招募了 200 名年满 18 岁居住在美国的研究参与人员,通过人口统计学调查、用户睡眠质量调查以及用户采用的策略等 3 部分调查,研究用户的睡眠行为对睡眠的影响以及用户会采用怎样的策略和工具提升睡眠质量。通过对问卷中开放性问题的定性分析得到用户的睡眠行为和改善的工具和策略,通过定量分析问卷中的结构化问题,发现年龄、职业和作息时间这些人口统计学特征以及相关的药物可能会影响睡眠质量,目前的睡眠改善

技术对于改善用户的睡眠质量没有显著效果。

2.23 环境和视觉素养(Environmental and Visual Literacy)

印第安纳大学的 Jacob Abbott 等采用在线调查和半结构化访谈方法调查了来自不同群体的 31 名参与者(环境科学博士生,建筑公司所有者,学生可持续发展委员会会议参加者、本科生),研究人们关于回收实践、资源保护习惯、技术使用、运输方法以及环境素养、对环境问题的态度等,旨在根据这些数据创建一个移动碳足迹计算器原型。研究表明大多数参与者有测量资源消耗及其对环境的影响的需求,认为测量结果的格式应该易读,并据此构建了碳足迹计算器原型 Creen,该计算器将利用来自物联网设备的用户输入和实时数据来测量能耗并输出个性化的碳足迹报告。普瑞特艺术学院的 Monica Maceli 等对 4 名档案实践工作者进行焦点小组访谈,焦点小组活动包括根据预定义的数据和仪表板功能列表进行投票、两人一组绘制仪表板的视觉外观,旨在探讨档案从业者在环境监测方面对数据仪表板的需求及挑战,提出基于用户需求和挑战初始数据的仪表板模型。研究表明档案实践工作者在使用数据仪表盘时面临技术、资源、物理空间、数据背景知识、交流、组织相关的挑战,技术和资源占主导地位。新加坡管理大学的 Yong Ying Joanne Tan 和 Tin Seong Kam 采用统计检验方法计算新加坡公共住房邮政编码的空间关联(LISA)的局部指标,并在 R 环境中进行分析,根据居住类型(1 室和 2 室公寓,3 室公寓,4 室公寓,5 室和行政公寓)进行可视化,探讨新加坡地区空间关联的局部指标(LISA)在电力消费数据中的应用,发现用电量异常值及高用电量的家庭的空间分布情况,识别当地的电力消费模式集群。

2.24 面向社会问题(Addressing Social Problems)

北卡罗来纳大学教堂山分校的 Yuanye Ma 以中国提出的社会信用体系为例,从文化视角探讨隐私问题,认为隐私是一种文化概念,其在中国文化中的意义尚未得到充分探索,并提出比较各国隐私文化的重要性。该研究作为概念性探索为未来的隐私工作开辟出新的研究领域。谢菲尔德大学的 Sally Sanger 等采用半结构式访谈方法,调查 25 名不同背景的用户在酒精在线支持小组论坛中的信息搜寻行为,采用 NVivo 11 编码数据,使用 Braun 和 Clarke 的专题分析方法开发主题。结果显示,使用酒精在线支持小组论坛形成饮酒问题想法的过程与 Bates 提出的“拾野莓”信息搜索模型相匹配,信息搜寻行为随着时间的推移而变化,用户使用多种来源和搜索技术来完善想法。南卡罗来纳大学哥伦比亚分校的 Kim M. Thompson 等采用内容分析方法,分析澳大利亚主要图书馆从 2018 年 1 月 22 日至 2 月 3 日的 96 个招聘广告,从能力、残疾、种族、语言、性别等方面探究了图书馆的多元化与包容性。结果显示,澳大利亚图书馆的大多数招聘广告相对中立且不具有明显的歧视性。

3 总结

iConference 作为 iSchool 的年会,其录用论文主题涵盖了当前信息科学领域理论研究和应用实践的最新成果,能代表当前学术研究的前沿,以下将从研究场景、研究方法、研究数据等角度对本次会议进行概要总结。

3.1 研究场景

数据科学及基于数据驱动的研究范式的变革和发展,促使图书馆与信息学科与其他学科进行融合,在新的研究场景下发现研究问题。本次会议中涉及的研究场景包括生物科学、天文学、地球科学、软件工程、智能个人助理、智能家居、数字游戏、医疗健康、移动支付、数字音乐、数字人文、交通、环境保护、消防、共享出行等,围绕这些场景中所面临的用户、技术和信息相关的问题展开研究。

3.2 研究方法

本次会议论文中主要用到的研究方法包括问卷调查法、半结构化访谈、焦点小组、实验观察法、扎根理论、主题分析法、统计分析法、Photovoice、社会网络分析、案例分析、机器学习、数据挖掘、元分析、眼动实验等方法,其中扎根理论、半结构化访谈、问卷调查方法是最主要的研究方法。

3.3 研究数据

本次会议超过 80% 的论文都是通过各种方法采集数据,通过编码、统计分析或文本分析方法发现数据背后所蕴含的特征和规律,因此研究数据的获取是研究中极为重要的过程。本次会议论文中用到的研究数据既包括从网络上各类应用进行抓取的数据,如社交媒体(如 Twitter、微博)数据、在线讨论社区(如 Reddit)数据、学术论文(WOS)数据、互联网电影资料库(IMDb)数据、用户评论数据(如 Amazon、Bestbuy)、产品介绍、用户手册等,也包括研究者通过结构化访谈、问卷调查方法所得到的数据。

参考文献:

［1］TAYLOR N G, CHRISTIAN-LAMB C, MARTIN M H. ,et al. Information in contemporary society, 14th International conference, iCon-

ference 2019, Washington, DC, USA, March 31 -April 3, 2019, proceedings[C/OL]. Washington: Springer, 2019 [2019 - 05 - 08]. https://doi.org/10.1007/978-3-030-15742-5_1.
[2] IDEALS. IConference 2019[EB/OL]. [2019 - 05 - 08]. https://www.ideals.illinois.edu/handle/2142/102118.
[3] NRC awards and honors dataset[EB/OL]. [2019 - 05 - 08]. http://sites.nationalacademies.org/pga/resdoc/pga_044718.

作者贡献说明:

巩洪村:资料收集与整理,论文撰写;

沈雨田:资料收集与整理,论文撰写;

曹高辉:论文选题,思路设计,论文撰写与修改。

"Inform, Include, Inspire"——A Summary of iConference 2019

Gong Hongcun[1] Shen Yutian[1] Cao Gaohui[1,2]
[1] School of Information Management, Central China Normal University, Wuhan 430079
[2] Center for Data Governance and Intelligent Decision, Central China Normal University, Wuhan 433079

Abstract: [**Purpose/significance**] iConference is an important international conference in information science. By combing and summarizing the accepted paper in iConference 2019, this paper aims to provide reference for future related search in information science. [**Method/process**] From the four dimensions of research method, research topic, research process and research finding, this paper summarized reviewed 77 accepted paper on 24 themes in iConference 2019. [**Result/conclusion**] As the annual conference of iSchool, iConference's topic covers the latest achievements of theoretical research and application in the field of information science. From the accepted paper, we find that the data-driven research paradigm provides an opportunity for the integration and development of information science and other disciplines, and promotes libraries and information disciplines to discover research problems in the new research scenarios; semi-structured interviews, questionnaires, grounded theory, statistical analysis and other qualitative and quantitative methods are the main research methods; research data acquisition is an extremely important research process.

Keywords: iConference　conference review　data-driven　information science